Mean Shift Clustering

Konstantinos G. Derpanis

August 15, 2005.

Mean shift represents a general non-parametric mode finding/clustering procedure. In contrast to the classic *K-means* clustering approach (Duda, Hart & Stork, 2001), there are no embedded assumptions on the shape of the distribution nor the number of modes/clusters. Mean shift was first proposed by Fukunaga and Hostetler (Fukunaga & Hostetler, 1975), later adapted by Cheng (Cheng, 1995) for the purpose of image analysis and more recently extended by Comaniciu, Meer and Ramesh to low-level vision problems, including, segmentation (Comaniciu & Meer, 2002), adaptive smoothing (Comaniciu & Meer, 2002) and tracking (Comaniciu, Ramesh & Meer, 2003).

The main idea behind mean shift is to treat the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local maxima or modes of the underlying distribution. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence. The stationary points of this procedure represent the modes of the distribution. Furthermore, the data points associated (at least approximately) with the same stationary point are considered members of the same cluster.

Next we review several of the technical details behind mean shift (for further details, see (Cheng, 1995; Comaniciu & Meer, 2002)).

Given n data points $\mathbf{x}_i \in \mathbb{R}^d$, the multivariate kernel density estimate using a radially symmetric kernel[1] (e.g., Epanechnikov and Gaussian kernels), $K(\mathbf{x})$, is given by,

$$\hat{f}_K = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right), \tag{1}$$

where h (termed the *bandwidth* parameter) defines the radius of kernel. The radially symmetric kernel is defined as,

$$K(\mathbf{x}) = c_k k(\|\mathbf{x}\|^2), \tag{2}$$

[1]Note that the mean shift procedure has been extended to anisotropic kernels (Wang, Thiesson, Xu & Cohen, 2004).

where c_k represents a normalization constant. Taking the gradient of the density estimator (1) and some further algebraic manipulation yields,

$$\nabla \hat{f}(\mathbf{x}) = \underbrace{\frac{2c_{k,d}}{nh^{d+2}}\left[\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)\right]}_{\text{term 1}} \underbrace{\left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x}\right]}_{\text{term 2}}, \tag{3}$$

where $g(x) = -k'(x)$ denotes the derivative of the selected kernel profile. The first term is proportional to the density estimate at $\mathbf{x}$ (computed with the kernel $G = c_g g(\|\mathbf{x}\|^2)$). The second term, called the *mean shift* vector, $\mathbf{m}$, points toward the direction of maximum increase in density and is proportional to the density gradient estimate at point $\mathbf{x}$ obtained with kernel K. The mean shift procedure for a given point $\mathbf{x}_i$ is as follows: (see Fig. 1):

1. Compute the mean shift vector $\mathbf{m}(\mathbf{x}_i^t)$.
2. Translate density estimation window: $\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{m}(\mathbf{x}_i^t)$.
3. Iterate steps 1. and 2. until convergence, i.e., $\nabla f(\mathbf{x}_i) = 0$.

For a proof of convergence, see (Comaniciu & Meer, 2002).

The most computationally expensive component of the mean shift procedure corresponds to identifying the neighbours of a point in space (as defined by the kernel and its bandwidth); this problem is known as *multidimensional range searching* in the computational geometry literature. This computation becomes unwieldly for high dimensional feature spaces. Proposed solutions to this problem include, embedding the mean shift procedure into a fine-to-coarse hierarchical bandwidth approach (DeMenthon & Megret, 2002) and employing approximate nearest-neighbour hashing-based search (Georgescu, Shimshoni & Meer, 2003).

Finally, a limitation of the standard mean shift procedure is that the value of the bandwidth parameter is unspecified. For representative solutions to the bandwidth selection problem, see (Comaniciu, Ramesh & Meer, 2001; Singh & Ahuja, 2003; Wang, Thiesson, Xu & Cohen, 2004).

References

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17(8)*, 790–799.

Comaniciu, D. & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24(5)*, 603–619.

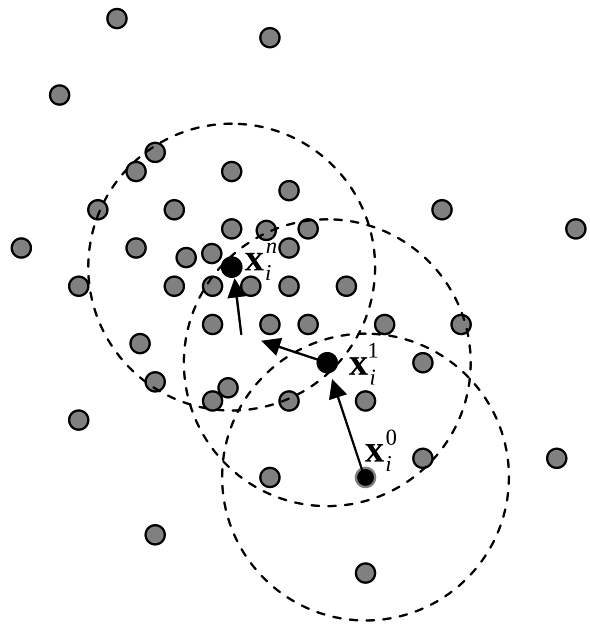

Figure 1: Mean shift procedure. Starting at data point $\mathbf{x}_i$, run the mean shift procedure to find the stationary points of the density function. Superscripts denote the mean shift iteration, the shaded and black dots denote the input data points and successive window centres, respectively, and the dotted circles denote the density estimation windows.

Comaniciu, D., Ramesh, V. & Meer, P. (2001). The variable bandwidth mean shift and data-driven scale selection. In *International Conference on Computer Vision* (pp. I: 438–445).

Comaniciu, D., Ramesh, V. & Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25(5)*, 564–577.

DeMenthon, D. & Megret, R. (2002). Spatio-temporal segmentation of video by hierarchical mean shift analysis. Technical report, University of Maryland.

Duda, R., Hart, P. & Stork, D. (2001). *Pattern Classification*. Wiley.

Fukunaga, K. & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, *21(1)*, 32–40.

Georgescu, B., Shimshoni, I. & Meer, P. (2003). Mean shift based clustering in high dimensions: A texture classification example. In *International Conference on Computer Vision* (pp. 456–463).

Singh, M. & Ahuja, N. (2003). Regression based bandwidth selection for segmentation using parzen windows. In *International Conference on Computer Vision* (pp. 2–9).

Wang, J., Thiesson, B., Xu, Y. & Cohen, M. (2004). Image and video segmentation by anisotropic kernel mean shift. In *European Conference on Computer Vision* (pp. Vol II: 238–249).